

UNITED STATES DISTRICT COURT**NORTHERN DISTRICT OF CALIFORNIA, OAKLAND DIVISION**

CHASOM BROWN, et al., on behalf of
themselves and all others similarly situated,

Plaintiffs,

vs.

GOOGLE LLC,

Defendant.

Case No. 4:20-cv-03664-YGR-SVK

PATRICK CALHOUN, et al. on behalf of
themselves and all others similarly situated,

Plaintiffs,

vs.

GOOGLE LLC,

Defendant.

Case No. 4:20-cv-05146-YGR-SVK

DECLARATION OF GLENN BERNTSON REGARDING**GOOGLE DATA SOURCES DISCUSSED AT AUGUST 4, 2022 HEARING**

1. My name is Glenn Berntson. I am an Engineering Director and Lead of the Google Ad Manager team at Google.

2. I am submitting this declaration to correct and clarify certain statements I made to the Court during the August 4, 2022 hearing on preservation issues. I attended that hearing on behalf of Google to help answer the Court's questions regarding technical aspects of Google's efforts to preserve data pursuant to the Court's orders. As described in further detail in the following paragraphs, in the course of designing the data preservation pipelines in [REDACTED] we have discovered that my understanding of the length of time for which data exists in [REDACTED], which I relayed to the Court during the August 4 hearing, was incorrect. Upon learning this, I immediately took steps to ensure

1 that Google will be able to comply with the Court's orders for data preservation. As described in
2 further detail in the following paragraphs, Google's ability to determine the length of time that data
3 will be stored in the [REDACTED] and [REDACTED] databases is subject to certain limitations, for example because
4 some types of data are not subject to time-based limits, but instead event-based limits. However, for
5 the data stored in [REDACTED] that is subject to time-based limits, the steps described below will ensure that,
6 consistent with the Court's Order, the majority of data in GAIA and DBL [REDACTED] present on July 30,
7 2022 will be retained until the data preservation pipelines are fully implemented.

8 3. In preparation for the August 4 hearing, I relied on my own knowledge of the [REDACTED] and
9 [REDACTED] storage spaces, which is extensive given that my teams rely on this infrastructure for
10 managing all of our user identity keyed data used in the serving of ads through Google Ad Manager
11 In addition to that, I also had conversations with members of the [REDACTED] and [REDACTED] team on other
12 details that I believed were relevant to the Court's determination.

13 4. During the August 4 hearing, to aid in the Court's evaluation of Google's proposed
14 timelines for implementation of [REDACTED] and [REDACTED] data preservation pipelines, I answered questions
15 from the Court regarding the length of time that data is stored in the [REDACTED] and [REDACTED] databases. In
16 particular, I stated that the data in [REDACTED] and [REDACTED] that is keyed off of user IDs can be stored for
17 different amounts of time, but the vast majority of it is stored for approximately [REDACTED]. I made
18 this statement based on my best understanding of the [REDACTED] and [REDACTED] databases, which is based on
19 my extensive experience supervising teams of engineers at Google who work with those databases.
20 I am also aware of data retention policies at Google which enforce a [REDACTED] lifetime of user IDs,
21 and the data associated with those IDs, for Ads purposes. I used that knowledge to draw an inference
22 regarding the amount of time that user ID keyed data will be stored in the [REDACTED] and [REDACTED] databases.

23 5. On August 18, I became aware that the [REDACTED] team had discovered that my statement
24 regarding the amount of data in [REDACTED] that is stored for [REDACTED] was incorrect. Specifically,
25 approximately [REDACTED] of GAIA-keyed data is stored in [REDACTED] for at least [REDACTED], and approximately
26 [REDACTED] of DBL-keyed data is stored in [REDACTED] for at least [REDACTED]. The percentage of GAIA- and DBL-
27 keyed data that is stored in [REDACTED] for at least [REDACTED] is approximately the same. The difference in the
28 data retention periods between the data keyed off of the different ID types is related to the observed

1 utility of the data over time. In particular, for GAIA, knowledge about a user's activity is useful to
2 inform ad serving decisions over a longer period of time than for Biscotti.

3 6. Upon learning that my understanding of the amount of user data retained over time
4 within [REDACTED] was incorrect, I immediately sought to confirm with the [REDACTED] team that my description
5 of the amount of time that data is stored in [REDACTED] was accurate. I confirmed that approximately
6 [REDACTED] of GAIA-keyed data is stored in [REDACTED] for at least [REDACTED], and approximately [REDACTED] of
7 GAIA-keyed data is stored in [REDACTED] for at least [REDACTED]. Approximately [REDACTED] of ZWBK-keyed data
8 is stored in [REDACTED] for at least [REDACTED], and approximately [REDACTED] of ZWBK-keyed data is stored
9 in [REDACTED] for at least [REDACTED]. This is consistent with my expectation and understanding when I
10 made the representation to the Court that the vast majority of data in [REDACTED] was stored for [REDACTED]
11 [REDACTED].

12 7. The calculations that I provide herein, for both [REDACTED] and [REDACTED], are based on an
13 analysis performed by the [REDACTED] and Kansas teams of the amount of data stored in columns that are
14 subject to a time-based limitation on storage of the data (commonly known as a [REDACTED]
15 [REDACTED]). Data may also be stored in columns subject to event-based limitations. In a column subject to
16 event-based limitations, such as [REDACTED] data stored in [REDACTED] and [REDACTED], for example, a field might
17 store the last 1,000 events and if that limit is reached the oldest event is discarded. While the [REDACTED]
18 and [REDACTED] teams were able to perform an analysis of the amount of time that data will be stored in
19 columns subject to time-based limitations, it was not possible to perform the same calculation to
20 estimate the amount of time that data is stored in columns subject to event-based limitations.

21 8. Following the [REDACTED] team's discovery of these issues, I promptly investigated the
22 amount of data stored in [REDACTED] that was at risk of not being preserved starting on July 30, 2022, due to
23 the fact that it has a shorter lifetime. I learned that a backup of the GAIA [REDACTED] and DBL [REDACTED] keyspaces
24 was created in the ordinary course of business on July 30, 2022. Google will retain this backup and
25 intermittent backups while the data preservation pipelines are in development. This will preserve
26 approximately [REDACTED] of the data that is subject to a time-based limit in GAIA [REDACTED] and DBL [REDACTED] from
27 July 30, 2022 until the data preservation pipelines are running, which is consistent with my statement
28 to the Court that the "vast majority" of this data would be retained for the same time period. These

1 backups will include columns that are subject to time-based limitations and columns that are subject
2 to event-based limitations. The backups will also include a much larger quantity of data than is
3 required under the Court's preservation orders.

4 9. The preservation order in Calhoun also requires Google to sample ZWBK [REDACTED]
5 However, the [REDACTED] team informed me that there is no longer a business purpose to collecting data in
6 ZWBK [REDACTED], and that only a small, residual amount of data is stored there. The team intends to
7 preserve a backup copy of all residual data in ZWBK [REDACTED] and cease collecting data in the keyspace.
8 Because ZWBK [REDACTED] will no longer collect data, ongoing daily sampling for that source would not
9 result in preservation of any data that is not already being preserved in the backup copy.

10
11 I declare under penalty of perjury of the laws of the United States that the foregoing is true
12 and correct. Executed in Jersey City, New Jersey on August 29, 2022.

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

DocuSigned by:
Glenn Berntson
048F0F988D984AA...

Glenn Berntson